False starts as modified output in second language learners of English: implications for proficiency

Introduction

This study reports a number of formal differences in the false starts produced by second language speakers of English at two levels of fluency. The results have implications for learning and teaching, testing and materials writing. Previous descriptions of false starts have relied on psycholinguistic taxonomies (Levelt, 1983; Kormos, 1998) based on their imputed function, and have resulted in definitional ambiguity and problems in application, e.g. failure to discriminate between those of different proficiency levels. In view of the importance of false starts as sites of language acquisition (Swain, 1998), it is clearly of interest to identify features that will enable finer discrimination to take place. The position taken is that such features can be discovered between the false start content produced by learners at contrasting fluency levels.

Method

Participants and procedures

The data for the study consisted of 56 speakers of English as a second language speaking 16 different first languages ($N_{females}$ = 36, M_{age} = 26.25, 9-45 years) who were, or were about to become, postgraduate students at a UK university. Participants were given one of six sets of semi-structured prompts based on Allen, Powell, and Dolby (2007) or Hashemi and Thomas (2011) at random and asked to produce 2 min continuous speech. Example topics included describing a business they would like to start, and something they had written that they were proud of. On the basis of auditing the speech samples produced by the participants, two EFL teachers assigned the participants to lower-intermediate or advanced categories, with reference to the public version of the Speaking Band Descriptors of the IELTS exam. An independent samples *t test* confirmed a significant difference, M_{low} = 5.17, SD_{low} = 0.17, M_{high} = 7.05, SD_{high} = 0.64, t (54) = 14.2 $p < .001$ between the proficiency levels (N_{low} = 25, N_{high} = 31). The samples produced were transcribed and coded. All false starts in the transcripts were identified by two independent raters and interrater reliability assessed by the two-way mixed, absolute agreement model to compute ICCs, which were within the accepted range ICC (3,2) = .94, $p < .001$. Any disagreements were re-evaluated and resolved.

In the present study, the false start data set (N = 167) was analysed by three pairs of expert raters in three iterations. The first two raters were asked to assign twenty randomised examples (ten from each proficiency group, pre-selected by the author for their concision and communicative content), e.g. 'we went [0.403][1] we .. we .. we hired a bus', into lower intermediate or advanced levels and afterwards to give reasons for their assignment. Their reasons included pausing, syntax and vocabulary. Pauses were excluded from further analysis, having already been shown not to be significant in this regard (Williams and Korko, 2019). A second pair of raters, both professional English language examiners, were given a

[1] Numbers in square brackets refer to pauses in seconds

similar task, this time to assign each example in the complete randomised data set (*N* = 167) to either a structural/superficial category, e.g. *my love .. favourite pet;* or to a conceptual category, e.g. *noticed it was .. actually [0.948^{pause}] so we couldn't see the stage*. The assignments were then compared with speaker level to note any association between descriptor and level. The results were not significant. In a third iteration, and following Wang et al (2012), who characterize text complexity as 'proportional to both the *syntactic* complexity and the *semantic* complexity' (Wang et al., 2012, p. 285), two new pairs of professional English language examiners assigned the data to one of three categories: *superficial*, i.e. a word search or minor modification to word form; *syntactic*, i.e. a revision of the phrase structure; or *conceptual*, i.e. the expression of a completely different idea (Table 2). These raters were also provided with approximately 50 words of the text surrounding the false start; and false starts relying on phonological cues for identification, and any occurring after the first in a compound sequence of false starts were removed from the set, leaving (*N* = 144).

Table 2. Example categories for the third rating iteration

False start category	**Example**
Superficial	I like watching football like Japan versus Korea <u>because . this is</u> because I can
Syntactic	and introduced our traditional <u>Chi.. culture of China</u>
Conceptual	<u>writing topic is [0.358] [1.227] I have</u> <u>OK I'm going to talk about</u> (a) business business I would like to start [0.526] and my ho.. [0.341] I'm going to run a business like a hotel

(False starts underlined).

Results

The interrater reliability for the third rater pair was found to be Kappa = 0.797 (p <.0.001), 95% CI (0.719, 0.875). The average similarity rate was 86.8%. A chi-square test of independence was performed to examine the relation between false start content and proficiency level. The relation between the variables was significant, X^2 (2, *N* = 144) = 6.655, *p* = .036. Adjusted residuals indicated that (1) advanced speakers were more likely to produce conceptual false starts and lower-intermediate speakers less likely (2.2, -2.2); and (2) lower-intermediate speakers were more likely to produce syntactic false starts and advanced speakers less likely (-2.0, 2.0).

Discussion

Levelt's (1989) speech production model suggests that the revision of an already conceptualised utterance involves more cognitive work than the accessing of grammatical structure or vocabulary *per se* as it is further removed from the moment of articulation and includes all three stages – conceptualiser, formulator and articulator (Wen, 2010). Advanced speakers are more likely to possess greater automaticity and can manage radical

reformulations (Skehan 2009). Lower-intermediate speakers are better able to handle syntactic revisions, which call for less cognitive reworking.

Implications

To maximise the production of modified output implies minimising demands on processing (Mackey et al., 2010) but motivating conceptual revision. Suitable tasks might therefore require information transformation (Skehan and Foster, 2001) and be relatively unstructured, though with essential lexis supplied. Such tasks are likely to elicit numbers of false starts with accompanying benefits for automaticity and acquisition.

References

Allen, M., Powell, D. and Dolby, D. (2007) *IELTS Graduation*. Oxford: Macmillan.

Hashemi, L. and Thomas, B. (2011) *IELTS trainer*. Cambridge: Cambridge University Press.

Kormos, J. (1998) A new psycholinguistic taxonomy of self-repairs in L2: A qualitative analysis with retrospection. *Even yearbook, ELITE SEAS working papers in linguistics*, 3, pp. 43-68.

Levelt, W. J. M. (1983) Monitoring and self-repair in speech. *Cognition*, 14, pp. 41–104.

Levelt, W. J. M. (1989) *Speaking: From intention to articulation*, Cambridge: MIT Press.

Mackey, A., Adams, R., Stafford, C. and Winke, P. (2010) Exploring the relationship between modified output and working memory capacity. *Language learning*, *60*(3), 501-533.

Skehan, P. (2009) Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied linguistics*, 30(4), pp. 510-532.

Skehan, P. and Foster, P. (2001) Cognition and tasks. In Robinson, P. (ed.) *Cognition and second language instruction* (Cambridge applied linguistics, pp. 183-205). Cambridge: Cambridge University Press.

Swain, M. (1998) 'Focus on form through conscious reflection.' In Doughty, C. and Williams, J. (eds.) *Focus-on-form in classroom second language acquisition*. Cambridge: Cambridge University Press.

Wang, Y., Berwick, R.C. and Luo, X. (2012) The cognitive complexity of texts and its formal measurement in cognitive linguistics. *Journal of advanced mathematics and computations,* 1(2), pp. 275-286.

Wen, Z. (2012) Working memory and second language learning. *International journal of applied linguistics*, *22*(1), pp. 1-22.

Williams, S. and Korko, M. (2019) 'Pause behaviour within reformulations and the proficiency level of second language learners of English.' *Applied psycholinguistics*, 40(3) pp. 723-742.